

INICIAR SESIÓN

NUESTROS PLANES

TODOS LOS
CURSOS

FORMACIONES

CURSOS

PARA
EMPRESAS

ARTÍCULOS DE TECNOLOGÍA

Histograma: qué es, ejemplos, gráficos y tipos

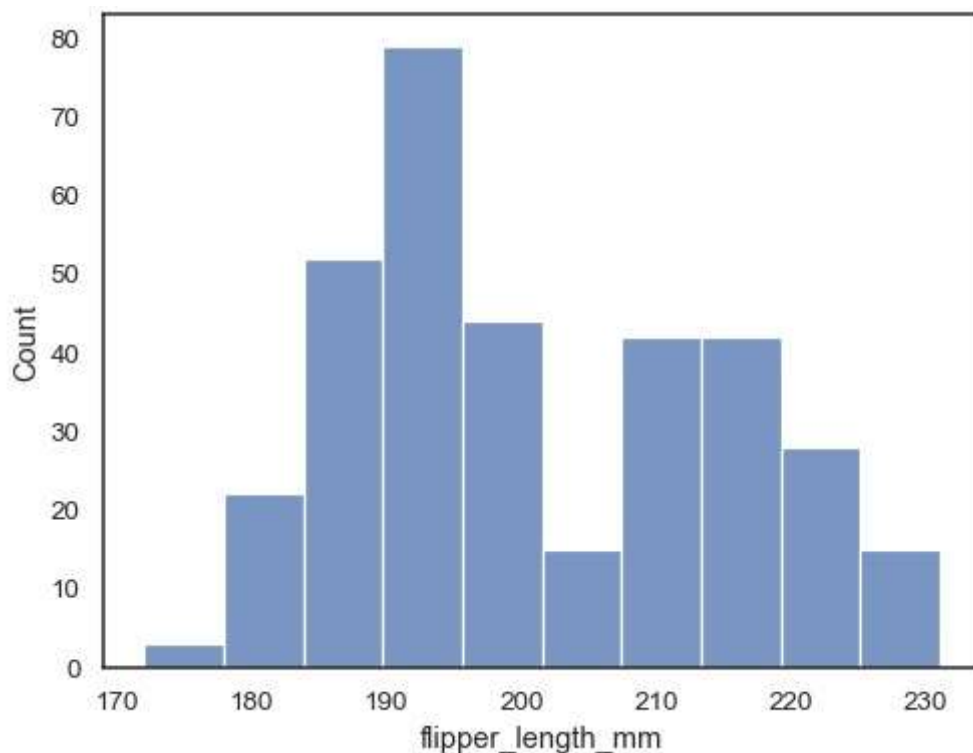


Daniel Siqueira

29 de Marzo



Un histograma es una especie de gráfico de barras que muestra una distribución de frecuencias. En el histograma, la base de cada una de las barras representa una clase y la altura representa la cantidad o frecuencia absoluta con la que ocurre el valor de cada clase. Al mismo tiempo, puede ser utilizado como indicador de dispersión del proceso. Este es un ejemplo de un histograma:



Cuando necesita presentar o sacar conclusiones de un gran conjunto de datos y está trabajando con conceptos que involucren frecuencias, ya sean absolutas o relativas, el histograma es la mejor manera de hacerlo. Nos ayuda con la representación gráfica de los conjuntos de datos de una manera más fácil de usar, lo que facilita ver dónde se concentran la mayoría de los valores.

Es útil crear un histograma cuando desea:

- **Resumir grandes conjuntos de datos visualmente:** a menudo, cuando utilizamos tablas, no es tan fácil sacar conclusiones. Podemos hacer nuestro trabajo más fácil y ganar mucho más tiempo y eficiencia usando un histograma.
- **Comparar los resultados:** es posible, con la ayuda del histograma, comparar rápidamente los resultados y, con la ayuda del eje y, saber, qué columnas excedieron los límites que necesitabas o no.
- **Comunicar la información de forma gráfica:** tanto los miembros de su equipo como los clientes pueden ver fácilmente los valores que ocurren con mayor frecuencia. Cuando utiliza un histograma para resumir grandes conjuntos de datos o para comparar resultados, está utilizando una poderosa herramienta de comunicación.

Una vez que hayamos recopilado los datos, el primer paso que vamos a dar es comprenderlos mejor, ya que nuestros cerebros pueden tener dificultades para

comprender un conjunto extenso de datos automáticamente. De esta manera, nuestra misión es hacer que la visualización de datos sea más inteligible y explícita.

Aquí es donde entra el histograma, ya que nos permitirá obtener la siguiente información sobre nuestro proceso:

- **Centralidad:** ¿cuál es el centro de distribución? ¿Dónde se espera que estén la mayoría de las observaciones?
- **Rango:** ¿la distribución normalmente contiene observaciones entre qué valores? ¿Cuál es el punto máximo y mínimo?
- **Simetría:** ¿debemos esperar la misma frecuencia de puntos de alto y bajo valor? ¿El proceso es simétrico o los valores más altos son más raros?

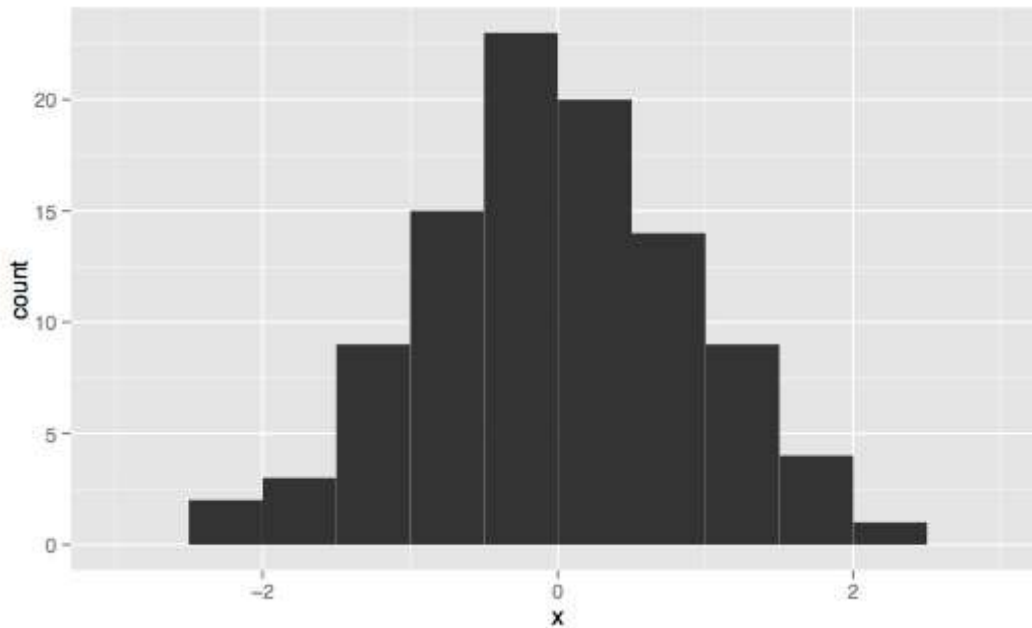
Los histogramas a veces se confunden con los gráficos de barras. Se utiliza un histograma para datos continuos, donde los intervalos de clase representan la extensión de los datos. Un gráfico de barras es un gráfico de variables categóricas o discretas. Algunos autores recomiendan que los gráficos de barras tengan espacios entre los rectángulos para aclarar la diferencia.

El propósito de un histograma es ilustrar cómo se distribuye una determinada muestra de datos o población, ordenando la información para facilitar la visualización de la distribución de datos. Al mismo tiempo, enfatiza la ubicación del valor central y la distribución de datos alrededor de este valor central. Ahora que hemos visto qué es un histograma, ¿serán siempre iguales, de un solo tipo? La respuesta es no. Tenemos diferentes tipos de histogramas, y conocerlos mejor puede ahorrarle tiempo y eficiencia en su análisis. Explore los tipos principales aquí.

Tipos de histograma

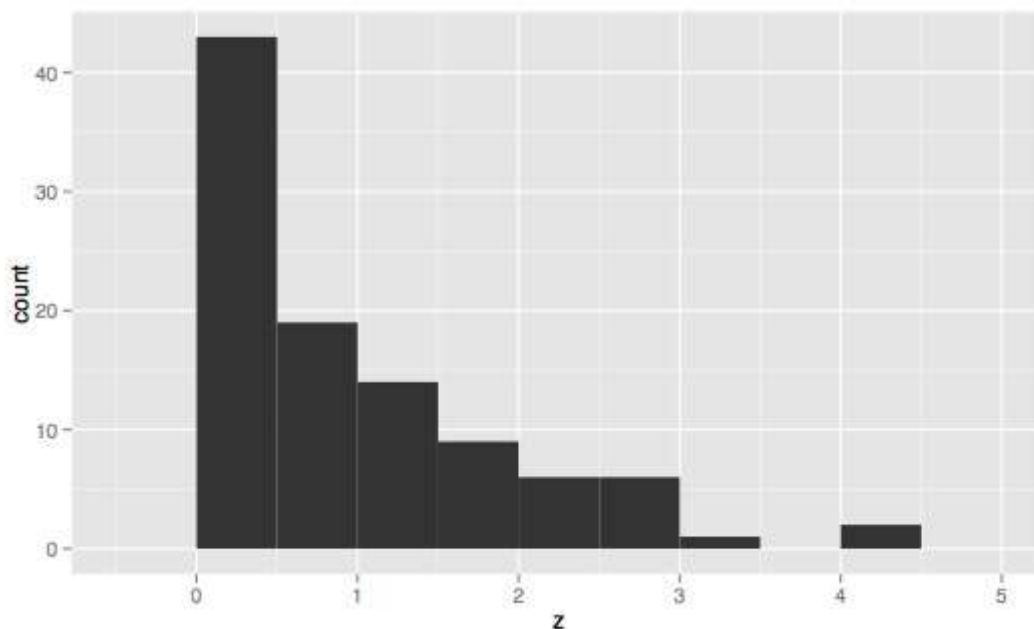
Simétrico

Un histograma simétrico (o unimodal) centra los datos en la media (medida central) y tiene características a través de la distribución de la media y la desviación estándar. Una característica del histograma simétrico es contener la mayor cantidad de datos del centro del gráfico. En estadística, este modelo se llama normal y le permite analizar cuánto se desvían otros datos de este modelo.



Derecha sesgada

Un histograma está sesgado hacia la derecha cuando la distribución de datos indica que los valores altos ocurren con baja frecuencia. Este modelo también se conoce comúnmente como el modelo de "cola derecha", ya que "se afina" a medida que nos movemos a lo largo del eje x, lo que indica que la frecuencia está disminuyendo.

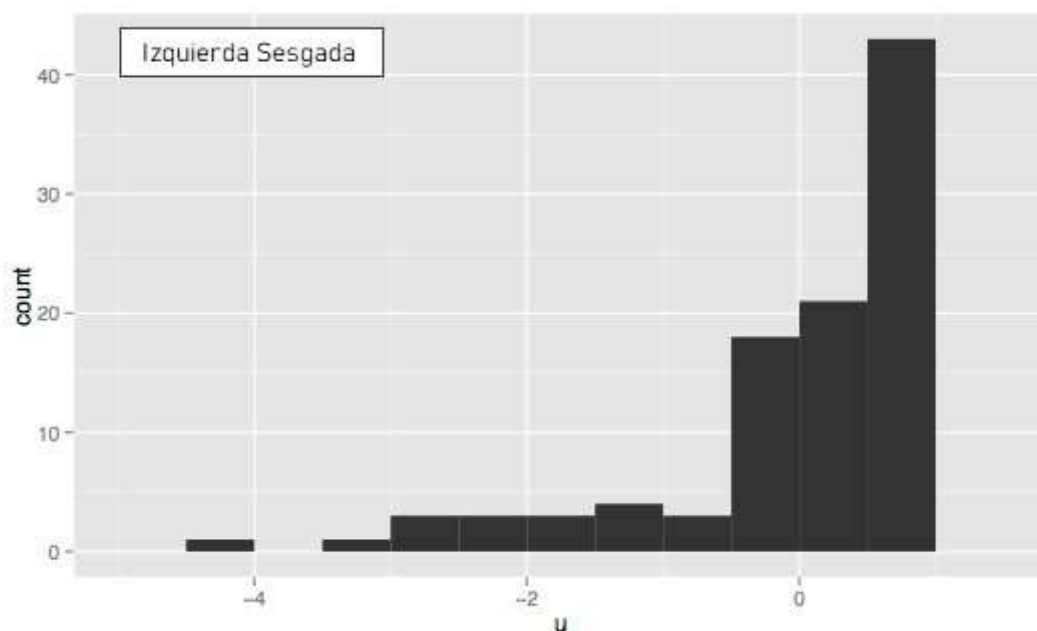


Por lo tanto, si te encuentras con un gráfico de este tipo, rápidamente podrás identificar el comportamiento de los datos.

Izquierda sesgada

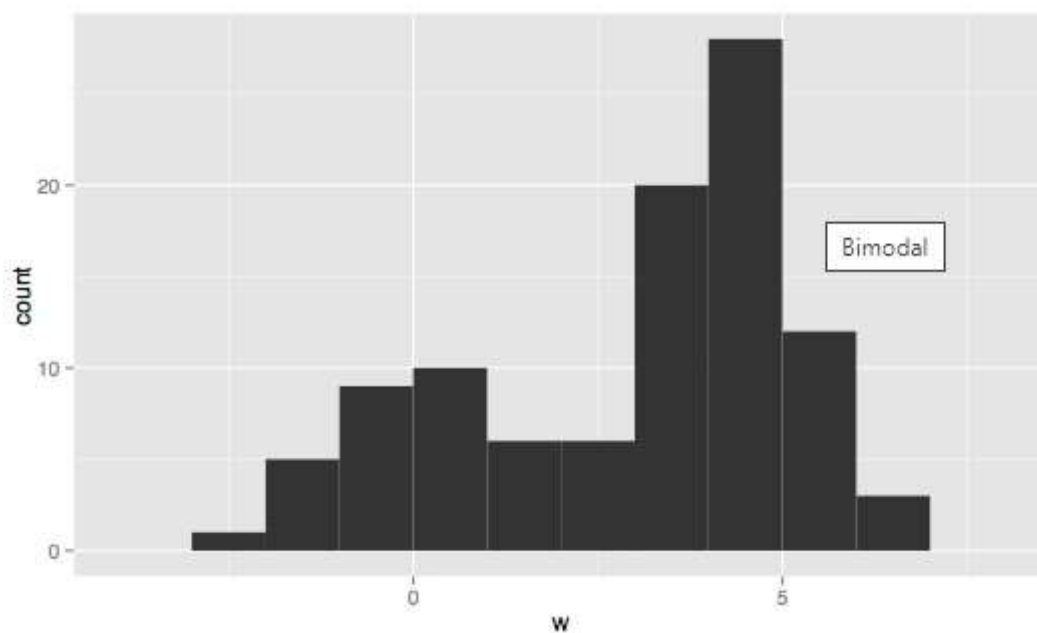
Llamamos de histograma sesgado a la izquierda cuando la frecuencia de los datos se concentra en los valores altos del lado izquierdo a medida que recorremos el eje x.

Podemos, por tanto, llamarlo también de histograma con “cola a la izquierda”, por la misma razón que el anterior, ya que a la izquierda formamos una especie de cola debido a la baja frecuencia de los datos al principio. Se observa que hay más información por encima de la media por falta de simetría.



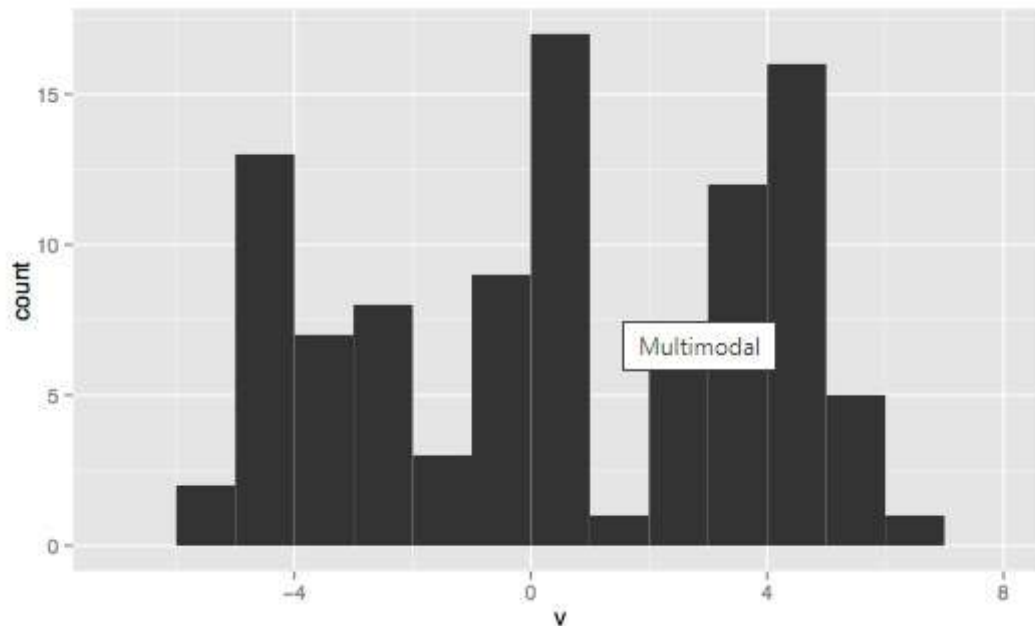
Bimodal

Llamemos al histograma bimodal cuando aparecen dos picos. Así sabemos que en dos momentos diferentes hay una concentración de frecuencia que sobresale.



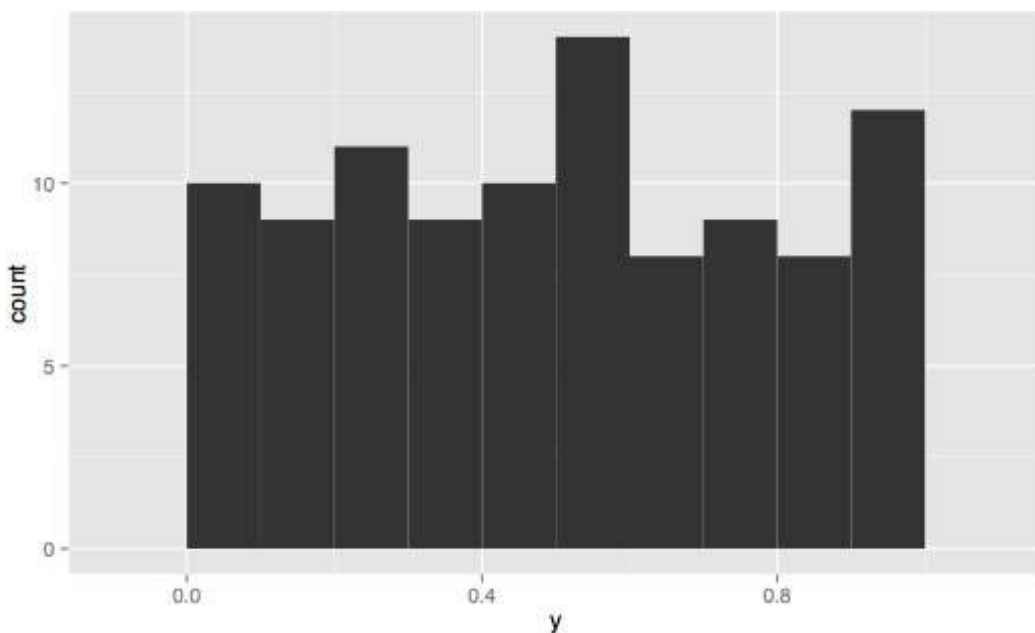
Multimodal

Un histograma es multimodal cuando aparecen varios picos. Los picos indicarán el mayor número de ocurrencias.



Meseta (aplanada)

Mucho se ha hablado sobre el “efecto meseta”. Esa palabra, “meseta”, nos recuerda un cierto tipo de aplanamiento, de igualdad constante de los datos. Un histograma tiene formato Meseta cuando sus barras tienen casi las mismas alturas. Esto ocurre cuando hay varias distribuciones juntas con diferentes medias.



Histograma en Python con Seaborn

Para trazar un histograma con Python, vamos a usar la biblioteca Pandas dentro de un notbook en Google Collab.

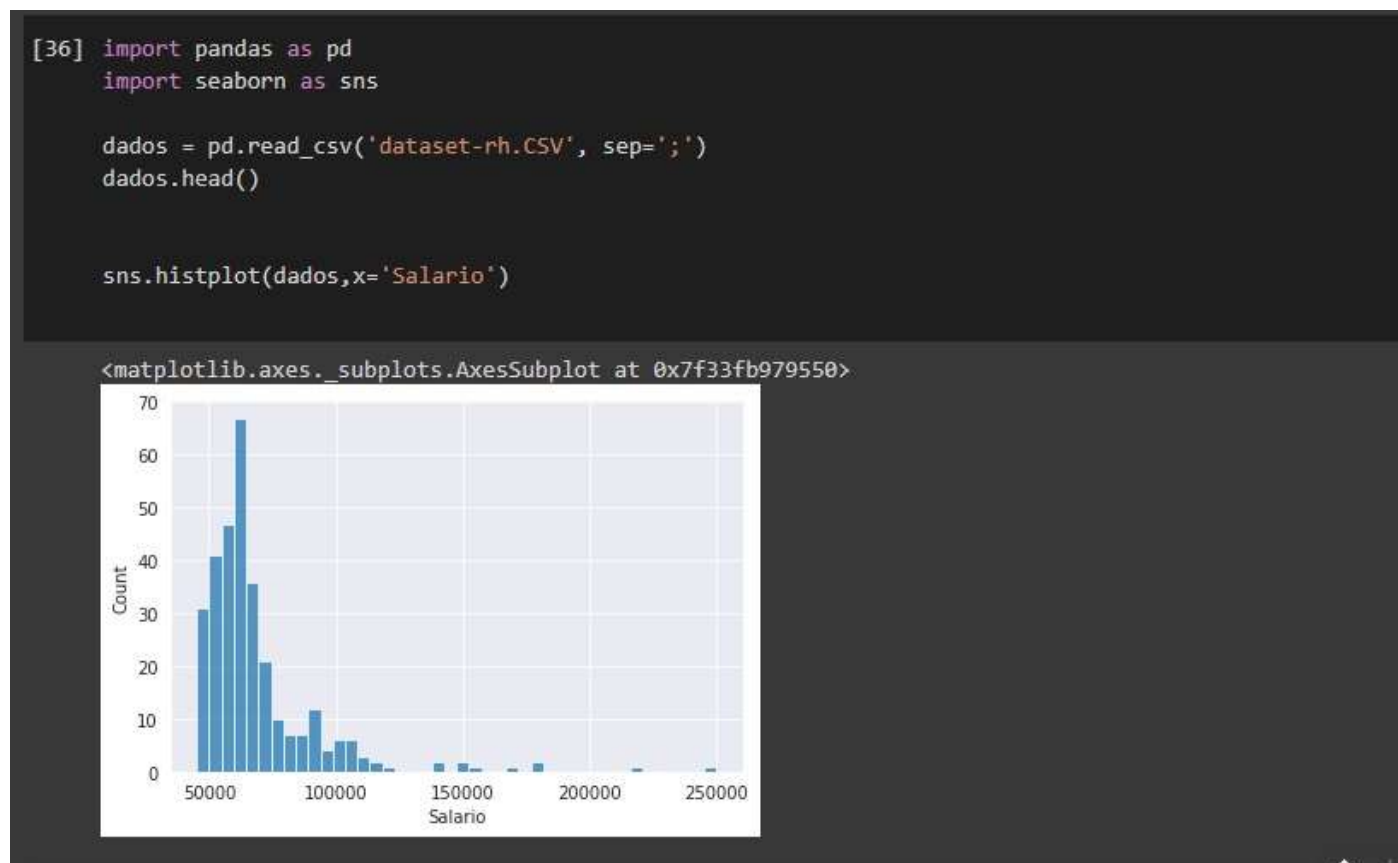
```
[9] import pandas as pd

dados = pd.read_csv('dataset-rh.CSV', sep=';')
dados.head()
```

	NomeEmpregado	IdEmpregado	Salario	Termd	Posicao	Estado	DataNascimento	Sexo	EstadoCivil	Nacionalidade	CorRaca	DataContratacao
0	Adinolfi, Wilson K	10026	62506	0	Técnico de produção I	SP	07/10/1983	M	Solteiro	Brasileiro	Branco	07/05/2011
1	Ait Sidi, Karthikeyan	10084	104437	1	Administrador de banco de dados sênior	SP	05/05/1975	M	Casado	Brasileiro	Branco	3/30/2011
2	Akinkuolie, Sarah	10196	64955	1	Técnico de produção II	SP	09/19/88	F	Casado	Brasileiro	Branco	07/05/2011
3	Alagbe, Trina	10088	64991	0	Técnico de produção I	SP	09/27/88	F	Casado	Brasileiro	Branco	01/07/2008
4	Anderson, Carol	10069	50825	1	Técnico de produção I	SP	09/08/1989	F	Divorciado	Brasileiro	Branco	07/11/2011

Imprimimos solo el encabezado, para una fácil visualización.

Ahora, tracemos el histograma para visualizar mejor la distribución de salarios en nuestra base de datos, utilizando otra biblioteca de Python conocida como seaborn. Tenemos la opción de usar el código `sns.histplot` para trabajar con el histograma.



solo mirando la imagen del histograma, podemos identificar rápidamente que está sesgada hacia la derecha (o con la cola hacia la derecha). Esto indica que la mayor parte de la concentración de datos se encuentra al comienzo de la distribución. Luego, analizando los datos, tenemos una vista rápida y fácil de la concentración de los salarios, que en su mayoría están entre \$ 50,000 y \$ 100,000 dólares anuales. Ahora podemos analizar visualmente el comportamiento de la distribución de frecuencias. Esto te ayudará a ganar tiempo y calidad en el análisis.



Daniel Siqueira

Daniel es instructor en la escuela de Datos y enseña Matemáticas, Física, Química e Inglés. Tiene una verdadera pasión por aprender cosas y temas nuevos, y transmitir sus conocimientos.

Traducido para Alura Latam por Daiana Righi

ARTÍCULOS DE TECNOLOGÍA

En Alura encontrarás variados cursos sobre . ¡Comienza ahora!

SEMESTRAL

US\$49,90

un solo pago de US\$49,90

- ✓ 218 cursos
- ✓ Videos y actividades 100% en Español
- ✓ Certificado de participación
- ✓ Estudia las 24 horas, los 7 días de la semana
- ✓ Foro y comunidad exclusiva para resolver tus dudas
- ✓ Acceso a todo el contenido de la plataforma por 6 meses

¡QUIERO EMPEZAR A ESTUDIAR!

[Paga en moneda local en los siguientes países](#)

ANUAL

US\$79,90

un solo pago de US\$79,90

- ✓ 218 cursos
- ✓ Videos y actividades 100% en Español
- ✓ Certificado de participación
- ✓ Estudia las 24 horas, los 7 días de la semana
- ✓ Foro y comunidad exclusiva para resolver tus dudas
- ✓ Acceso a todo el contenido de la plataforma por 12 meses

¡QUIERO EMPEZAR A ESTUDIAR!

[Paga en moneda local en los siguientes países](#)

Acceso a todos
los cursos

Estudia las 24 horas,
dónde y cuándo quieras

Nuevos cursos
cada semana

NAVEGACIÓN

PLANES

INSTRUCTORES

BLOG

POLÍTICA DE PRIVACIDAD

TÉRMINOS DE USO

SOBRE NOSOTROS

PREGUNTAS FRECUENTES

¡CONTÁCTANOS!

¡QUIERO ENTRAR EN CONTACTO!

BLOG

PROGRAMACIÓN

FRONT END

DATA SCIENCE

INNOVACIÓN Y GESTIÓN

DEVOPS

AOVS Sistemas de Informática S.A

CNPJ 05.555.382/0001-33

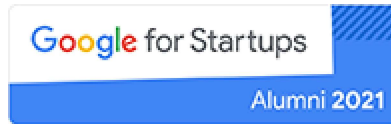
SÍGUENOS EN NUESTRAS REDES SOCIALES



ALIADOS



En Alura somos unas de las Scale-Ups seleccionadas por Endeavor, programa de aceleración de las empresas que más crecen en el país.



Fuimos unas de las 7 startups seleccionadas por Google For Startups en participar del programa Growth Academy en 2021

POWERED BY

CURSOS

Cursos de Programación

Lógica de Programación | Java

Cursos de Front End

HTML y CSS | JavaScript | React

Cursos de Data Science

Data Science | Machine Learning | Excel | Base de Datos | Data Visualization | Estadística

Cursos de DevOps

Docker | Linux

Cursos de Innovación y Gestión

Productividad y Calidad de Vida | Transformación Ágil | Marketing Analytics | Liderazgo y Gestión de Equipos | Startups y Emprendimiento